

DOCUMENT RESUME

ED 183 610

TM 800 016

AUTHOR Strenio, Andrew, Jr.
TITLE The Debate over Open versus Secure Testing: a
Critical Review. National Consortium on Testing Staff
Circular No. 6.
INSTITUTION Huron Inst., Cambridge, Mass.
PUB DATE Oct 79
NOTE 76p.
EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Answer Keys; *College Entrance Examinations; Cost
Effectiveness; *Disclosure; Educational Legislation;
*Educational Testing; Federal Legislation; Higher
Education; Public Opinion; Standardized Tests; State
Legislation; *Testing Problems; *Test Results
IDENTIFIERS *Open Testing; Testing Industry; *Test Security;
Truth in Testing Legislation

ABSTRACT

The context of the debate over secure versus open testing is described, with particular emphasis on tests used for postsecondary or professional school admission. Recently introduced federal and state legislation on testing is also described. Four arguments in support of open testing are summarized: (1) fairness demands that tests be open; (2) open testing will increase the testing industry's accountability; (3) it will increase the public's knowledge of testing and lead to improved tests; and (4) test takers will learn from the information supplied to them. Next, the three main arguments for secure testing are reviewed: protecting test quality; minimizing the costs of testing; and minimizing the problems of teaching to the test and the proliferation of coaching schools. Following an analysis of each of these arguments, comments are made concerning the political, judicial, and situational factors (related to the type of test and impact of testing procedures) which help to clarify the broader perspectives on this debate. It is concluded that, although open testing may have unanticipated and undesirable consequences, it seems that the current status of secure testing needs to be changed. Questions are raised concerning the best means of accomplishing this goal. (GDC)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

The Huron Institute

NATIONAL CONSORTIUM ON TESTING PROJECT

123 MOUNT AUBURN STREET
CAMBRIDGE, MASSACHUSETTS 02138
617-681-9285

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

ED183610

NATIONAL CONSORTIUM ON TESTING

STAFF CIRCULAR NO. 6

THE DEBATE OVER OPEN VERSUS

SECURE TESTING: A CRITICAL REVIEW

October 1979

Andrew Strenio, Jr.
The Huron Institute
123 Mount Auburn Street
Cambridge, Ma. 02138

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Paul L. Houts

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

This is the sixth of a series of staff circulars distributed as part of the Huron Institute's staff work for the National Consortium on Testing. This circular is intended to spur discussion on the topic of open testing and to raise important issues worthy of additional attention, not to resolve them. It is produced primarily as a resource document for member organizations of the NCT. Although the circular has been prepared as part of staff work for the Consortium, it does not necessarily reflect the views of the Consortium nor of individual Consortium members.

IMS00 018

ACKNOWLEDGMENTS

The author would like to thank the following people for their helpful comments during the preparation of this staff circular: Nancy Burton, Emily Cahan, David Cohen, William Corbett, Peter Cowden, Jack DeSanctis, Richard Ferguson, Mike Garet, Alice Irby, Tom Kinsock, Mark McQuillan, George Madaus, Les May, Fritz Mosher, Allan Nairn, Gene Radwin, Gene Russianof, Jack Schmidt, Laurie Scott, Stephen Solomon, John Weiss, David White, Elinor Woods, Mark Yudof, and Jerrold Zacharias. Other people who took the time to make their views known on the subject and deserve a word of thanks include: Wellington Grimes, Stanley Kaplan, Senator Kenneth LaValle and staff, Arthur Levine, Robert Moulthrop, Jean Nazzaro, Frances Quinto, Vito Perrone and Roger Shuy. A special thanks for their invaluable assistance and support goes to my colleagues at the Huron Institute, especially Norma Robbins, Barbara van Fossen, Maria Sachs, and Mariellen McNicholas. The author is indebted as well to any other contributors who have been inadvertently omitted, and to Judith Strenio for her tolerance and criticism. Walt Haney, most of all, through his editing talents and generosity with his time and friendship, deserves credit for improving this paper.

None of these people, of course, is to be blamed for any errors of fact or interpretation that remain in the paper. In particular, tentative conclusions suggested do not necessarily reflect those of anyone other than myself.

ABSTRACT

This report explores the debate over open versus secure testing. Introductory remarks explain the dynamics behind the increased public interest in this debate and define the types of tests that are being considered. The paper next presents the key arguments to be made for open testing and for secure testing and then reviews each proposition closely. Additional perspectives which do not fall neatly into either camp are then introduced. These factors are reviewed in an effort to get at the larger questions involved in the choice of a testing format, questions which tend to be overlooked in the debate over open versus secure testing.

The report proceeds to weigh the arguments, both pro and con, and to reach some tentative conclusions regarding the relative merits of open and secure testing in general. New directions for this evolving policy are discussed in light of the need for particular types of information.

TABLE OF CONTENTS

	<u>Page</u>
Introduction	1
I. Context of the Debate	3
II. Exposition of Arguments	9
A] The Case for Open Testing	9
2.1 Fairness	9
2.2 Accountability	10
2.3 Information for dialogue and better tests	13
2.4 Educational value for the tested	16
B] The Case for Secure Testing	17
2.5 Preserve test quality	17
2.6 Lower cost	19
2.7 Lesser test preparation	21
III. Analysis of Arguments	23
IV. Broader Perspectives	45
V. Conclusion	54
Footnotes	63

INTRODUCTION

The debate over open versus secure testing involves many controversial issues. Are current secure standardized selection tests fair when those tested cannot review the questions and intended answers that may determine their educational opportunities? Would open testing diminish test validity or make current equating procedures impossible? Do test publishers need to be regulated to increase their accountability to the public? How much more would tests cost under open testing? Is the battle over open and secure tests really a proxy war involving larger issues of testing policy? Where will -- where should -- this debate lead us?

These questions, and many others closely connected to them, are not simple to answer. This report explores the implications inherent in the questions, some different approaches to them, and ends with some tentative conclusions. The structure of this paper is straightforward. Part I describes the context of the debate and provides a few initial definitions and remarks. Part II then presents the case for change made by proponents of open testing and the case for the maintenance of present practices by defenders of secure testing. The arguments are recounted largely in the terms used by participants in the debate. Part III then attempts to sift through the issues raised, to separate fact from opinion, to probe weaknesses and assess strengths. Part IV takes a step back from all the trees of specific points under contention to view the forest of related testing issues. Part V sums up the competing arguments and conveys the author's tentative conclusions, in addition to suggesting constructive approaches for future work.

So far much of the public debate on this topic has been notably disjointed. Opposing sides tend to argue past each other without really joining fire on central points. The interplay -- or lack of it -- between the two sides in some ways resembles a brief encounter in a revolving door. One person speeds from the street into the revolving door and back into the street; while the other person rushes from the building into and around the door and back into the building. There is a blur of action, but both parties end up largely where they started having gone through a confined space without quite touching.

This lack of direct encounter by opposite parties to the debate makes it difficult, as we will see, to definitively resolve many of the arguments at issue. Moreover, the problem is further complicated by the continuing nature of the debate which makes it difficult to pin down arguments for review and analysis. For example, while there has been much testimony on open testing in legislative hearings, official transcripts of these debates are not yet available. However, I do offer tentative conclusions regarding some of the issues raised. I have attempted to be fair to the various perspectives represented in the debate and have sought objectivity even though I do not claim impartiality. Readers should, therefore, be warned of my potential biases. My effort to separate arguments from an evaluation of them will not be considered successful by all. Indeed, one reader of a draft version of the paper charged that the body of the text was entirely biased, while another interested party complained that the conclusion was too neutral. Such perceptions of partiality and neutrality doubtless always reside to a certain extent in the eyes and interests of the beholder, so the reader is left to render an independent judgment.

PART I: CONTEXT OF THE DEBATE

The debate between defenders of currently secure tests and proponents of open testing has recently grown to national prominence. Advocates of open testing argue that in the interests of fairness and testing industry accountability the questions and answers used on standardized tests should be made public shortly after test administration. Advocates of secure testing argue that disclosure of tests is not feasible for several reasons including potential invalidation and increased cost. While this debate has historical roots -- Walter Lippmann railed against "cults of secrecy" in testing as long ago as 1922 -- it has intensified greatly in the past 2 years. This increasing prominence may be due to several factors: for example, increasing use of tests generally and widespread debate about their appropriate use and consequences. With baby boom children competing for advanced schooling and jobs, the consequences of scores achieved on secure tests have become more severe. One critic of secure tests, Jon Haber, writes:

Standardized tests exist everywhere in our society. They are used to decide who will go to our colleges and universities, who will get advanced placement credit, who will qualify for financial aid, and ultimately who can be professionally licensed after their training.¹

With so much seemingly at stake, more and more students and parents want to know exactly what questions are asked on the tests.

Whatever their exact cause, these demands have gained momentum from legislative initiatives made by critics of current testing practices, who have taken their arguments to the United States Congress and to state

legislatures (notably in California and New York).² Generally, the call has been for passage of "truth in testing" laws which would require test publishers to release test questions and answers -- as well as statistical information -- soon after test administration. Supporters of secure testing have been quick to oppose such bills, stating that the current limited disclosure helps guarantee test quality while open testing would invalidate the use of secure tests and necessitate new and more costly test development programs.

The "truth in testing" movement has thus far resulted in new legislation in both California and New York. In California, Governor Brown signed testing legislation (SB 2005) into law on September 26, 1978, effective January 1, 1979. Under that law, sponsors of tests required for admission to post-secondary educational institutions must deposit copies of tests equivalent to those actually used with the California Postsecondary Education Commission. These facsimiles will then be available "for possible inspection by scholars in the field."

SB 2005 was passed only after it was amended to allow deposit of facsimiles rather than tests actually in use. Since these facsimiles may be old tests, or equivalents designed solely for deposit with the Commission, some open testing advocates are dissatisfied. Stephen Solomon, a member of the New York Public Interest Research Group (NYPIRG) and a leader of its open testing project, feels that SB 2005 is worse than no law at all. He maintains that the law doesn't require test sponsors or publishers to disclose anything vital while allowing them to claim they are regulated by the state.³ As an alternative, NYPIRG supported a bill with stricter provisions in the New York legislature. Governor Carey

signed this bill into law on July 14, 1979, effective January 1, 1980. The New York law requires: (1) full disclosure of questions and answers determining examinees' scores on post-secondary and professional school admissions tests within thirty days of the release of test scores; (2) special statistical reports about test validity to be filed by publishers; (3) background information on the tests to be mailed to all test takers; and (4) specific authorization by test takers before agencies disclose test scores to anyone other than designated recipients. In response to the New York law, the American Association of Medical Colleges (AAMC) and the Association of American Dental Colleges announced on July 17, 1979 that they would stop giving their medical and dental school admission examinations in New York State rather than comply with the "truth in testing" law.⁴

At least partly because of the controversy in New York, consideration of national testing legislation has begun in the U. S. Congress. Hearings were held before the House Subcommittee on Elementary, Secondary, and Vocational Education to consider two such legislative proposals: the "Truth in Testing Act of 1979" (H.R. 3564) sponsored by Congressman Gibbons of Florida, and the "Educational Testing Act of 1979" (H.R. 4949) sponsored by Congressman Weiss and Congresswoman Chisholm of New York, and Congressman Miller of California.⁵

In addition to these proposed federal actions, the 1975 Education for All Handicapped Children Act (P.L. 94-142) seems to bear on the general question of disclosure. According to Garry McDaniels, a division director in the Bureau of Education for the Handicapped, the Act provides

. . . an opportunity for the parents of a handicapped child to examine all relevant records with respect to the identification, evaluation, and educational placement of [the] child . . . and to obtain an independent educational evaluation of the child.⁶

In McDaniels' view, "If the parents and guardians wish to challenge the interpretations of the available information, they are welcome to do so; the work of the measurement community is open to inspection and challenge."⁷ The argument that may be anticipated is that all parents and children should have an equal right to inspect test results and challenge them whether children are labelled handicapped or not. The argument does not necessarily follow, however, since Congress itself treated the handicapped as a special group.

In short, the truth in testing legislation in New York and California,⁸ legislative proposals now before the U.S. Congress, and implementation of the Education for All Handicapped Children Act have all served to intensify the debate over open versus secure testing. Before describing arguments under debate, a few prefatory remarks are in order. First, the truth in testing legislative proposals have several distinct elements. One is that test publishers or sponsors file information on test validity, reliability, and cost with state or federal agencies. Another element is disclosure of test content to test takers after administration -- called open testing in this paper. Other elements are release of certain information to prospective test takers, and guarantees of confidentiality for individuals' test scores. While all these elements are important, this paper deals mainly with the open testing proposals -- both because they are likely to have the most far-reaching consequences, and because they have been the center of the most heated debate.

Second, "open" and "secure" are used in this paper as their advocates use them. Open testing refers to disclosure -- either through publication,

return of corrected results or unlimited access to files -- of the questions and answers on a test after that test has been given. Secure testing refers to the current practice of some test agencies of keeping actual test questions and answers secret even after administration. For the sake of brevity, the terms "open testers" and "secure testers" will be used to refer to the supporters of open and secure testing respectively.

As a matter of semantics, many advocates of open testing prefer the term secret rather than secure to label their opposition. Ending secrecy seems to be a stronger rallying cry than reducing security. Strictly speaking, secure tests are indeed secret -- in the sense that they are "kept from public knowledge or from the knowledge of a certain person or persons."⁹ But the term is pejorative. For this reason, and on the principle that proponents of particular viewpoints are entitled to choose their own nom de guerre, this paper will use the more neutral term.

Third, although arguments reviewed in this paper are couched in fairly broad terms, the tests in contention are of mainly one type: secure norm-referenced multiple-choice tests used in selection decisions. The arguments both in favor of and against the secure nature of these tests thus may be only partly relevant to other types and uses of tests.¹⁰ It should be stressed that the New York law, which is the basis for much of the discussion of open testing in this paper, is specifically limited to tests "designed for use in the process of post-secondary or professional school admissions."¹¹ Tests in this category include those used for

admission to college (American College Test program, ACT; and the Scholastic Aptitude Test, SAT); graduate school (Graduate Record Examination, GRE; Law School Admissions Test, LSAT; Graduate Management Aptitude Test, GMAT; and the Medical College Admissions Test, MCAT); and certain professions (bar and medical board exams). With respect to testing for purposes other than admission, some of the arguments reviewed in this paper also take on a different coloration. Some of these perspectives and the varying hues they may lend to the open versus secure testing debate will be reviewed in the closing section of this paper.

PART II. EXPOSITION OF ARGUMENTS

What are the views in dispute in the open versus secure testing debate? This section reviews the essential points of contention, advancing with little elaboration the arguments in favor of both open and secure testing.

A] The Case for Open Testing

The main arguments in support of open testing boil down to four propositions which may be paraphrased as follows:

- 1) Fairness demands that tests be open;
- 2) Open testing will lead to increased accountability on the part of the testing industry;
- 3) Open testing will result in greater public knowledge of testing, and to improved tests; and
- 4) Test takers will learn from information supplied by open testing.

2.1 As a matter of simple fairness, a person is entitled to know in detail the basis upon which he or she is judged. The first argument for open testing translates into an appeal for fairness in the use of tests as instruments of public policy. Test results have important effects upon the lives of the tested. If test takers are expected to abide by the results with equanimity, they should be able to examine and evaluate the test contents. Tests are seen as vehicles for ranking people. With secure tests, those ranked have little opportunity to understand, much less to challenge, why the rankings come out as they do. This theme is prevalent in the writings of open testing supporters. A selection of quotations making the fairness point follows. Edwin Taylor of the Education Development Center (EDC) argues:

- Standardized tests, those given to large numbers of people, are major instruments of public policy. They influence decisions about who is promoted from grade to grade, who is put into special classes, who is admitted to colleges and universities, who enters law schools and medical schools, and who is selected for various administrative positions. These instruments of public policy are not available for public examination and discussion.¹²

Test critic John Shlien claims:

- If these decisions importantly affect our private and communal lives, it is important for the public to know a good deal about the tests. The public does not. In many instances, the public is not supposed to.¹³

William Corbett, principal of the James Russell Lowell School in Watertown, Massachusetts, argues:

- When . . . "competency" . . . is under scrutiny, we believe it is a fundamental if not a civil right to know the basis on which that "competency" is to be determined.¹⁴

Expositions of the fairness argument have occasionally cited the 1977

Ethical Standards of Psychologists, which states:

Persons examined have the right to know the results, the interpretations made, and, where appropriate, the original data on which the final judgments were based. Test users avoid imparting unnecessary information which would compromise test security, but they provide requested information that explains the basis for decisions that may adversely affect that person or that person's dependents.

This statement obviously reflects the fairness theme, but just as obviously leaves open the question whether disclosure is "appropriate" when tests are used for various sorts of selection or licensing.¹⁵

2.2 Open testing is needed to help curtail the monopoly power of test companies and to make them more accountable to the public. Advocates of open testing want to make test companies and sponsors more accountable to the public for the quality and consequences of the tests they develop and administer. State Senator John Dunlap, sponsor of the California legislation, expresses this view:

ETS [the Educational Testing Service, one of the nation's largest testing organizations] is accountable to no outside or public agency and the monopolistic power it wields over the lives of millions and the revenues and political power it generates for itself shows a private industry accountable to nobody but itself.¹⁶

In speaking of a "monopoly" open testers seem to imply that the testing industry has little effective competition and can therefore engage in over-pricing and callousness toward consumers. Open testers see government regulations requiring test disclosure as the first step in remedying this situation. They believe public scrutiny of test content will restrain testers by placing them under the scrutiny of an aroused citizenry.

In support of their claim of industry "non-accountability" open testers have compiled a list of recent "errors" in the construction and scoring of tests and the reporting of results. For example, according to several accounts, ETS changed the content of the LSAT between the July and October 1977 administrations without notifying the nation's law schools. The result was that nearly twice as many students scored 750 or better in October than in July. Further, the number of perfect scores (i.e., 800s) was three times as high as normal. Law schools with rolling admissions policies accepted students on the erroneous assumption that the October LSAT scores meant the same thing as those from July 1977 and earlier. This, it was charged, reduced the chances for admission of students who took the LSAT before October 1977. According to these accounts, ETS did not explain the score variance to law schools until February 1978.¹⁷

In another incident, ACT, which administers the MCAT, informed medical schools in May 1978 that the MCAT scores sent to them were incorrect.

Due to an error in the scoring key sent to ACT by the MCAT development agency, roughly 90 percent of the scores were too low.¹⁸ Other incidents cited were a 10 point scoring discrepancy on the GMAT;¹⁹ and the incorrect designation of students taking the LSAT as "unacknowledged repeaters." In effect, it was said, law schools were told erroneously that certain candidates were certified liars. Critics charged that even after the error was discovered, ETS made no effort to inform the alleged "unacknowledged repeaters" that they had been so designated.²⁰

Some advocates of open testing see such incidents, which they label errors, as the inevitable result of secrecy in testing, since under current practices test takers cannot know whether mistakes have been made, let alone rectify them. William Corbett dismisses standardized secure test results as "numerical scores . . . determined by a machine subject to mechanical error, fed by attendants subject to human error."²¹ Test critic John Williams suggests that test takers have little chance to challenge the power of test publishers:

What happens when a person taking a test recognizes an error in the instrument? If the experience of the students reported by [Banesh] Hoffmann is typical, writing to a test company is of little avail. Apparently, they rarely respond to individuals writing to question certain items -- at least they rarely admit to error.²²

Some critics claim they need access to test content in order to properly evaluate the need for test reform or even test abolition. While not explicitly impugning the integrity of test publishers, many open testers argue that current procedures are heavily weighted in favor of the publishers' self-interest.²³ Other critics, concerned about possible

cultural bias, ambiguity, and the like in the tests, claim that non-accountability may have pernicious results, and that test security hinders a careful and impartial examination of these issues. One unusual but true example of a stifled attempt at discussion is cited below:

Three students in a graduate school of education take a national standardized test for elementary school principals. They are horrified by many of the questions, which they think are ambiguous, poorly phrased, implicitly racist, and inappropriate in content for certification of elementary school principals. The students write an article about this test for a national magazine, quoting twelve test questions from a copy obtained for them by the editor of the magazine. Before publication they send the article to the company that developed the test. The company threatens to sue the editor if the article is published since, they say, he violated an agreement of secrecy entered into when the test was sent to him. The article is not published.²⁴

In sum, open testers maintain that the public cannot evaluate the criticisms made of tests when the questions themselves are perpetually hidden from scrutiny.²⁵

2.3 Open testing will be beneficial in leading to increased knowledge about tests and improved tests through a public dialogue. Open testers say that not enough information reaches the public about the content of secure standardized tests and what they can and cannot do. The truth in testing laws, according to their sponsors, are efforts to increase the flow of information from a trickle to a torrent. NYPIRG claims: "... a misunderstanding of what tests can do and not do has contributed to a rising dissatisfaction with schools. This legislation is a first step toward giving educators, parents and students the material they need to make more informed decisions."²⁶ In California, the student sponsors of SB 2005

claimed:

... the release of validity studies and the test questions and answers will give researchers, who are independent from the testing companies, an opportunity to compile in-depth information on the test.

We realize that the testing companies have long performed their own sort of evaluation studies. But this is much like asking a student to grade his or her own paper; naturally the student's ability to judge fairly will be distorted by the fact that it is to his/her advantage to grade her/himself highly.²⁷

Open testers also claim that test security results in misunderstanding of test results. Not only may scores be misinterpreted by unsophisticated users who don't allow for standard error and coaching/practice effects, but the basis for test results may be misunderstood. Roger Shuy of the Center for Applied Linguistics suggests: "Discrimination, unexplained, is empty education, in my opinion. I want to know why a test question discriminates, not just that it does."²⁸ Another case cited by Edwin Taylor of EDC reflects this contention:²⁹

In a suburb of Boston, Massachusetts, a parent is told that his daughter has performed poorly on a standardized test of English and therefore will be placed in a "special learning" class. He requests that a copy of the test with his daughter's answers be put into her school files so that he can examine them. The request is denied, first by the school principal, next by the superintendent of schools, and then by the elected school board that is legally responsible for the local system.

The parent forces the school board to grant his request by threatening to sue under the Massachusetts Freedom of Information Act. Finally, he and his daughter look at the test together. Her difficulty is not with the test but with the separated answer form. For example, with a red pencil she easily corrects errors of punctuation and capitalization in a sample paragraph but makes mistakes in coding the corrections and transferring them to the answer sheet. For her, the original test was not at all a test of English but a test of clerical skill.³⁰

An open test arguably would have revealed the "clerical skill" problem sooner and with less aggravation for all concerned. Although this example does not come from the realm of post-secondary tests, the difficult task of distinguishing test taking skills from the more fundamental skills ostensibly measured is relevant at all levels of testing.

Open testing, according to supporters, would stimulate an informed discussion of the ends and means of testing. When it comes to deciding on testing policies with profound consequences for individuals and society, open testers feel the public has a right to participate directly and not merely through "professional surrogates." At least one testing expert, Robert Ebel, writing in 1963, seems to indirectly support that view:

But test makers have, I fear, sometimes given the general public reason to fear that we may be up to no good. I refer to our sometime reluctance to take the layman fully into our confidence, to share fully with him all our information about his test scores, the test from which they were derived, and our interpretations of what they mean.

Secrecy concerning educational tests and test scores has been justified on several grounds. One is that the information is simply too complex for untrained minds to grasp. Now it is true that some pretty elaborate theories can be built around our testing processes. It is also true that we can perform some very fancy statistical manipulations with the scores they yield. But the essential information revealed by the scores on most educational tests is not particularly complex. If we understand it ourselves, we can communicate it clearly to most laymen without serious difficulty. To be quite candid, we are not all that much brighter than they are, much as we may sometimes need the reassurance of thinking so.

. . . If specialists in education measurement want to be properly understood and trusted by the public they serve, they will do well to shun secrecy and to share with the public as much as it is interested in knowing about the methods they use, the knowledge they gain, and the interpretations they make. This is clearly the trend of opinion in examining boards and public education authorities. Let us do what we can to reinforce the trend. Whatever mental measurements are so esoteric or so dangerous socially that they must be shrouded in secrecy probably should not be made in the first place.³¹

While Ebel's remarks were directed mainly at test score disclosure, they do seem relevant to the current debate.

Advocates contend that open testing would dispel the "testing mystique" that is a side effect of testing secrecy. Public discussion and examination of tests, they suggest, would lead to better tests, used and interpreted with more accuracy. Oscar Buros, long-time publisher of the Mental Measurements Yearbook, suggested that more information on secure tests is needed:

I would like to recommend that every secure test be accompanied by a manual describing the planning, construction, standardization, reliability, and validity of the test. Since the tests themselves can be examined only under highly restricted conditions, if at all, it becomes of greatest importance that detailed information be provided for secure tests. The important role which these tests play in influencing the lives of the examinees is much greater than the role of commercially available tests. Yet far less information is available on these secure tests and, to make matters worse, much of the information which the publisher does possess is considered to be confidential--for in-house use only. If detailed information were made available simultaneously with the publication of tests, not only would educators and testing specialists be in a better position to evaluate the tests and to interpret the results more intelligently, but the publishers would be impelled to construct tests, pointing out their strengths and weaknesses.³²

Most open testers, however, are not prepared to settle for more "detailed information" from test publishers in lieu of the tests themselves.

2.4 Open testing will enable examinees to learn from their corrected test results. Some open testers hold that open testing could enhance the value of testing as a learning experience³³-- that if examinees receive corrected versions of their test (or an answer sheet) soon after taking the test, they can go over incorrect answers, figure out why they were incorrect, and avoid repeating their mistakes. In this respect, open testers assert that test secrecy can disguise the task before the examinee.

Though not speaking strictly of post-secondary admissions tests, Lazarus and Taylor make this point in a more general fashion;

The publishers say that the tests must be secret so children cannot prepare for them. But if the tests were good, preparing for them would be a fine idea. Everyone knows the range of questions that will be on the driver's license tests, and by practicing and preparing for them, it is possible to become a better driver. Children could become better thinkers in much the same way if they were given good, open tests.³⁴

In sum, proponents of open testing think that security conceals inequities and inadequacies in tests from the light of public exposure. They maintain that open testing would help ameliorate many of the problems of secure testing. Their arguments for change concern fairness to the tested, accountability of the testing industry, test improvement through public dialogue, and enhanced individual learning. What are the counter-arguments in favor of maintaining the status quo with respect to secure testing?

B] The Case for Secure Testing

The main arguments in support of secure testing seem to boil down to three propositions:

- 1) Secure testing is necessary to protect test quality;
- 2) Secure testing helps to minimize the costs of testing; and
- 3) Secure testing minimizes problems of teaching to the test.

2.5 Secure testing is necessary to preserve test quality in terms of validity and equating. In essence, secure testers claim that releasing test questions would greatly endanger both current methods of validating standardized tests and also the statistical equating of test results over

time (so that a score of say 625 on the SAT means today what it meant fifteen years ago). Since secure standardized tests now re-use questions in successive test administrations, full disclosure of this year's test would either preclude re-use of questions (ending equating procedures) or lead to prior exposure to questions (threatening tests' validity).

The point is made in several ways. For example, the College Entrance Examination Board (CEEB) claims: "Once the questions and answers of any test are non-secure, the value and fairness of using that particular form of the test for critical educational decisions is lost."³⁵ Moreover, CEEB argues that open testing would disrupt "... technical procedures through which tests are standardized which, in this case, means constructing tests so that they will be reliable and consistent from test to test in their level of difficulty and their powers to assess whatever is being measured."³⁶ ETS claims: "Disclosure would also make impossible the existing process of equating scores on the same scale from one year to the next Pretesting new questions would also be impossible under the [open testing] requirements."³⁷ In a minor variation on this theme, the Associated Medical Schools of New York argues: "Since there is common agreement that a finite limit exists on the number of good questions which could be written ... disclosure requirements would have an increasingly negative effect on the validity of the exam as more and more questions were disclosed."³⁸

Secure testers, then, claim that if the high quality of secure tests is to be preserved, open testing would compel them to make every form of

every test entirely new at each administration. The College Board put it this way:

Whenever standardized tests are used to help make critical decisions relating to college admission, financial aid, and credit purposes, they need to be highly secure. Once the questions and answers of any secure test are disclosed, the value and fairness of re-using that particular form of the test or any of the questions on it is lost, and a new form of the test must be constructed.³⁹

2.6 Open testing would require new and expensive methods of test development in order to preserve test quality. Secure testers claim that open testing would greatly increase costs which would have to be passed on to the consumer. First, there would be the added cost of developing new questions (if question re-use was to be avoided) and the cost of pretesting and equating alternative test forms. The College Board has written: "Our test development costs are now relatively small because we can re-use test questions we keep secure. And our test scores are now a uniform national academic standard because we can repeat 'equating' sections over time."⁴⁰ ETS points to the elaborate process used to develop questions in the quest for validity, reliability, and cost effectiveness:

Test questions are written to sets of specifications developed for each test. Contrary to myth, there is no vast storehouse or computer full of appropriate secure test questions waiting to be placed on a test. Questions must be written and tests assembled by individuals with training and experience in test development as well as appropriate subject fields. It is not, for example, possible to use the same question for both the SAT and the LSAT. In large measure, different questions must be written for different purposes on different tests.

The test development process for a single question will typically include editorial reviews, reviews by appropriate external groups, reviews by minority groups, pretesting, and final editing. Questions written by individuals at ETS and across the nation for any nationally standardized examination may go through as many as 100 different steps over a period of 24

months before appearing on an actual test. The test development process thus involves a large measure of interaction between those individuals who put the test together, and those individuals who work with external bodies--the College Board or the Law School Admission Council, for instance--who determine ultimately the content and the specification for the test. Disclosure of test questions would thus add an enormous additional burden on the current test development process that would affect cost and service to students.⁴¹

Second, there would be additional real, though indirect, costs to the educational system. Secure testers argue that the number of times a certain test is offered might be reduced, creating hardships for those unable to take it on a particular occasion. ETS states: "Reduction in the number of tests available in [New York] would distinctly limit the possibility of special test administrations for special reasons. . . ."⁴² Regarding the disclosure provision of the New York law, ETS argues:

The primary effect of this section must be either to reduce the number of times a test can be offered within the state, or to create the need for the development of a large number of special forms of every national test for use in the state.⁴³

Secure testers have given a range of estimated cost increases likely to stem from open testing. According to the College Board, the cost of the SAT, now \$8.25, would go up by about \$4 per student per test. CEEB's estimate, made in response to proposed legislation in California and New York, is based on developing 20 new forms a year at \$40,000 per form, or \$800,000.⁴⁴ In a memorandum opposing the NYPIRG bill, the Associated Medical Schools of New York claims an even more drastic impact on the cost of the MCATs for New York students. The MCAT is given four times a year in New York to a total of 4000 takers. At an estimated cost of \$25,000 to develop a new MCAT form, the additional cost of \$100,000 would be divided among the 4000 takers, so that the price per student would rise by \$25. Moreover, the Association

warns that some takers may pay more, such as the 300 to 500 people who take the test on Sundays, who face a possible cost increase of \$100.⁴⁵

Secure testers thus argue that if open testing becomes the rule test takers will have to foot substantial additional direct costs, as well as indirect costs in the guise of fewer opportunities to take particular tests and diminished test quality.

2.7 Open testing would exacerbate problems of test preparation and lead to a proliferation of cram schools. While its supporters contend that open testing would bring into perspective the strengths and weaknesses of tests for selection purposes, secure testers hold that the opposite is likely to occur. They envision testing running amuck as teachers and students, knowing the questions and answers to previous tests, devote more and more time to learning the test instead of the underlying subject matter. In this scenario, students would flock to cram schools (or published guides) charging exorbitant rates for minimal returns.

Three statements are representative of this belief. According to the 'Associated Medical Schools of New York, " . . . the availability of test questions and answers to commercial review course enterprises would create a bonanza for them without any commensurate value to the student."⁴⁶ More comprehensively, ACT has written:

Book stores and newspaper ads provide ample evidence of the many entrepreneurs who publish previously used test items and provide 'cram' courses to gullible students who happen to be able to pay the typically exorbitant fee charged. As evidenced by our own research on repeat testing, the results are predictably disappointing. We believe the proposed legislation plays right into the hands of these entrepreneurs. They, rather than students and institutions, would be the real benefactors [sic] of the legislation.⁴⁷

ETS contributes the following: "A loss of [test] confidentiality could give an unfair advantage to the few who could avail themselves in advance of copies of test questions and answers."⁴⁸

In sum, proponents of secure testing claim that open testing is a Pandora's box, not a panacea; and that the current secure tests, whatever their flaws, are preferable to the situation that open testing would create.⁴⁹ ETS succinctly states its arguments against the New York legislation: "In sum, the legislation would require duplicative reports, increasingly costly test development procedures, a concomitant loss of service to candidates, increasing fees for tests, and a possible diminution in the quality of the tests themselves."⁵⁰ Secure testing advocates believe that the shortcomings of open tests, coupled with the relatively satisfactory performance of secure tests, add up to a persuasive argument that society should not precipitously do away with current security in post-secondary admissions and selection tests.

PART III: ANALYSIS OF ARGUMENTS

The above arguments seem to be the key elements in the debate over testing. What are we to make of them? Let us begin to make sense of the debate by examining critically each of the seven arguments cited above.

3.1 The Fairness Argument

One observer of the debate, law professor Mark Yudof, has criticized the fairness argument in the following terms:

Why is open testing equated with fairness? Can't a secure test be fair? Is a closed hearing necessarily unfair? And besides, one needs a concept of fairness or justice to make these judgments. Fairness may lie in utility (identifying those with merit), distributing goods and services to people based on other characteristics (e.g., need), or achieving a random distribution of success and failure on tests among cognizable groups, e.g., racial minorities, women, white males. I suspect that the question is more one of legitimacy. We do not ask whether a neurosurgical technique is fair; for it gains legitimacy through expertise.⁵¹

It seems axiomatic that secure tests and closed hearings can be fair in fact.⁵² Indeed, opponents of open testing, quoted above, have cited fairness as a reason for maintaining test security. But secure testers' references to fairness deal mainly with test quality and test preparation. These aspects of the fairness issue are discussed below in sections 3.5 and 3.7. Nonetheless, the fairness argument does seem to be the strongest weapon in the open testing arsenal. As Walt Haney has suggested:

The main argument in favor of open testing is the proposition that as a matter of simple justice, individual test-takers have a right to know the exact manner in which they are judged, and the details of the assessments which may⁵³ affect their educational opportunities and life chances.

The proposition has a certain appeal, but the basis for it is unclear. Would the argument be extended, for example, to the claim that applicants have a right to know the exact basis on which they were refused admission to an institution?

How might we go about deciding whether currently secure tests are fair? One might propose some sort of systematic review of all secure tests. This would doubtless be a formidable undertaking, but even apart from that there is ample question of its likely outcome. Research on test bias suggests that reviewers tend to disagree on which test items are potentially biased against particular minorities or cultural groups.⁵⁴ Yet test publishers assure us that minority groups regularly review items for bias. What are we to make of this contrast?

At a minimum, it raises the question of the symbolic fairness of secure tests. Secure standardized tests have been under sustained attack from many sources. While these attacks may be ill-informed, they have at least called into question the apparent fairness of secure tests for all manner of minority and distinctive groups. Yudof sees this as a challenge not to the objective fairness of tests, but to their perceived legitimacy. The point, he writes:

. . . may be that tests are not perceived as legitimate because we no longer trust the experts to eliminate bias, and this is reinforced by the actual outcomes of testing. Open-testing may be a form of popular participation in test formation which would legitimize it in the eyes of open-testers. Or they might ultimately insist on legitimacy only when designated outcomes occur ⁵⁵

If testing experts are indeed no longer widely trusted, or if the general mistrust of authority in American society indeed now extends to encompass

the ranks of testing professionals, then it is hard to see how anything less than public disclosure of the inner workings of testing can overcome public cynicism about tests. This is not to say that cynicism would necessarily be reduced by disclosure -- indeed, under certain scenarios cynicism might even increase. However, release of test questions and answers would at least undercut the argument that "if tests are secret it must be because someone is hiding something." This seems to be the gist of the argument that Thomas O. White, Vice-President of Law Programs for ETS, made in a memorandum to the Law School Admission Council dated July 24, 1979. White wrote:

. . . I am also convinced that until we set it out for legislators, critics and the general public to see, we will never carry the day with any of them. Few will accept that the imagined simply is not so. Rumor and inference will continue to prevail over fact and we run the risk of losing the free use of the very tools with which much of our success has been fashioned.

. . . our credibility and the quality of our program can be regained and maintained only if we move, as quickly as possible, to full disclosure of all aspects of the Law School Admission Services Programs. . . . Our programs should have no reason to be concerned about critical scrutiny. . . . We need only to broaden our critical audience to assure that this matter does not become a political contest focused on our motivation instead of the substance of our services.⁵⁶

Yet even if the symbolic fairness of tests could be assured with respect to all the parties to the current debate, it would still leave unresolved the actual fairness of particular tests used in particular ways with particular individuals. The point here is that fairness resides not so much in any one test as in the use of that test for particular purposes. If we view standardized tests as social instruments for

accomplishing specific purposes, asking whether or not a test is fair is like asking whether a hundred-yard dash is a fair means of selecting professional football players. For certain positions, and other things being equal, it may be fair; in other circumstances, probably not. This line of reasoning suggests that, whatever value open testing may have in increasing the fairness of testing practices, it cannot be expected to completely eliminate unfairness or misuse of test results with respect to particular individuals.

The fairness argument, then, could be viewed as calling for a cost-benefit determination. Would any harm to the value of the tests themselves (i.e., to their reliability, validity, cost effectiveness) be outweighed by gains in public and individual confidence in the essential fairness of the testing process? My judgment here is largely influenced by one of Yudof's comments. He wrote:

I think that more should be made of the secrecy (as opposed to the security) point. All things being equal--the arguments on both sides seem weak to me--we have a preference for government in the open. Testing is a form of governance, and a form without other accountability constraints. For example, I think of the recent executive order requiring the routine declassification of secret documents when no strong case can be made for continued secrecy. The documents may reveal nothing, but democratic governments should err on the side of openness.⁵⁷

The premise behind this line of reasoning is that testing is a form of governance, that post-secondary admissions tests are instruments of public policy. For tests which millions of students each year are required to take the premise seems plausible, but how far it could be extended seems problematic. Insofar as the most widely used post-secondary admissions tests do seem to have grown into de facto instruments of public policy, the call for open testing as an aspect of open governance does indeed seem warranted.

In sum then, whether seen as a matter of individual justice, as a vehicle for increasing public confidence in tests, or simply as a taste for open proceedings in governance -- fairness as advocated by open testers is a strong argument for open testing.

3.2 The Accountability Argument

Test publishers argue with some justification that they are already accountable to their governing bodies, and through them to the public. For example, ETS must answer to the Law School Admission Council, a body of the nation's law schools, for its work in developing LSAT and to other such bodies for its work on other tests. ACT maintains that it "... is governed and controlled by representatives of the institutions and agencies which use its services."⁵⁸ While some have questioned the degree of independence of the governed from the governing, it is true that the overseeing bodies provide at least some check on the testers.

Moreover, the College Board has also raised important questions regarding the ownership of some of the material open testers want disclosed. During the discussion of the California law, CEEB claimed that requirements for submitting data on predictive validity to the state bureaucracy were "... addressed to the wrong parties. Validity studies and data on which they are based belong to individual institutions, not to test organizations."⁵⁹ Some test publishers have also argued that open testing legislation may illegally infringe their proprietary rights to their tests.

The accountability argument is not susceptible to proof one way or the other, given the scant evidence on this topic. Rather, it appears to derive strength mainly from the increasingly general clamor for greater

control over all sorts of official bodies -- be they legislatures (through meetings open to the public), law enforcement agencies (FBI guidelines), intelligence agencies (CIA Congressional committees), or energy producers and utilities (cost and supply disclosure requirements). Still, the national fervor for increased accountability of all public institutions should be tempered by exploration of the need for change in any given sphere. Mere faddishness would not be a sufficient rationale for tampering with the testing system.

There are two aspects of the accountability argument which, if accepted, would seem to transcend charges of "change for change's sake." The first is that open scrutiny of tests would lead to fewer errors in individual scores, since test takers would comb their results for poor questions or scoring errors that would lower their scores. The scoring and reporting "errors" cited earlier indicate that improvement in this regard is possible, but its extent is quite unclear. Moreover, test publishers view the uncovering of errors not as evidence of the need for new means of accountability, but as indications of the effectiveness of present safeguards, and they view the potential for improvement in this realm as quite small.⁶⁰ Such disclaimers, of course, are not uncommon for organizations reluctant to become the object of government regulation.

The second accountability tenet of open testers is more general. It is that open testing would directly enfranchise test takers. Currently, the review bodies to whom test publishers are accountable are sponsoring organizations -- such as the LSAC and the American Association of Medical Colleges (AAMC), to name but two. Test takers pay their money, take the tests, gain or lose by their scores, and yet have

no institutional representation of their interests, nor a ready process for error correction or challenges independent of test publishers and sponsors.⁶¹ The interests of the tested may diverge from the interests of the test sponsors and publishers at various times, and the sponsors and publishers are not accountable to the tested (or to society as a whole, for that matter) in the way that publishers are accountable to sponsors. But the real implications of this argument are problematic. At the symbolic level, the return of corrected test results to test takers would certainly seem to increase testers' accountability to test takers. However, students receive the corrected results of teacher-made tests, without necessarily having much recourse to challenge dumb questions or capricious grading.⁶²

To the extent, then, that one believes in accountability as an end in itself, open testing may produce an incremental improvement of the status quo, at least symbolically. Would scoring errors be reduced? Possibly, although the size of the reduction and its importance in the overall scheme of things eludes precise estimates. Would publishers be more directly accountable to test takers? Yes, at least at the symbolic level. So this argument lends some weight to the case for open testing, though the practical consequences of such accountability are unclear.

3.3 The Public Knowledge and Test Improvement Argument

It is difficult in our society to argue in favor of withholding information from the public except in circumstances of national security or personal privacy, and open testers have taken advantage of that fact. New York State Senator Kenneth LaValle, chairman of the Higher Education Committee and one of the sponsors of the New York law, asserted: "This is just an

information bill. We are asking that the companies that prepare these tests give us a little more information about them."⁶³ Yet in discussing the reporting requirements of a proposed federal bill (H.R. 4949), Walt Haney interjected a note of caution:

Some of these proposed requirements strike me as somewhat ironic since it is fairly clear that substantial misuse of standardized tests over the last decade has not simply grown out of the practices or policies of educational institutions or test publishers but has been prompted in significant measure by reporting requirements of federal and state governments.⁶⁴

Nevertheless, at least some test publishers acknowledge the legitimacy of demands for increased disclosure of test information. Robert Solomon, ETS executive vice-president, has observed that ETS is ". . . very much in sympathy in a general way with the objectives that the originators of the [New York law] seek to serve." He points out that in recent years ETS has given students substantially more information on test purposes, scoring and margins of error; and that in the last two years, potential test takers have been provided with free sample copies of the SAT.⁶⁵ The controversy in this area, then, centers on how much new information should be publicly disseminated, and how soon.

In terms of the open testing provisions of legislation, the point in contention is what would be gained if test takers and the public generally had access not just to facsimile test versions, but to the contents or corrected results of tests actually taken. It might be argued that test takers will pay more attention to actual than to facsimile test questions, but it is hard to see how this would necessarily increase public knowledge about testing generally. The effect of numerous past efforts to educate the public about standardized testing has been uncertain.

Moreover, the point often overlooked in the commonly acknowledged notion that it would be a good thing for people to know more about standardized testing is that different parties have some very different ideas about what the public should know.⁶⁶

If access to actual test contents did not increase public knowledge about testing, might it improve test quality itself? Test publishers typically recount that review of test items by diverse people (measurement experts, substantive experts, and minority group members) is a useful part of test development. Perhaps wider reviews by test takers, and their objections to specific test items, would lead in time to improved test items. But three points are worth noting. First, as previously stated, when different persons review test items from particular perspectives, they tend to identify very different sets of items as potentially defective. Second, the success of such an evolutionary process would doubtless depend on the establishment of mechanisms for test publishers to receive and adjudicate test takers' complaints about specific items. Third, whether tests would improve through open testing would very likely depend on how we judged test quality. By some standards, test quality might improve, but by others it might decline.⁶⁷

In sum, while in principle open testing might create a more knowledgeable public, any real increase in public comprehension of tests generally would seem unlikely unless open testing were supplemented with other efforts aimed at increasing public understanding. Any improvement in test quality itself would depend on several factors including how one went about assessing such quality. The added information might make " . . . the debate [over testing] . . . less civil and more acrimonious," in the words of Yudof. He goes on

to speculate:

Not being familiar with the theory of test construction, critics are likely to dwell on individual questions that strike them as wrongly answered, sexist, racist, etc. I would still argue, however, that open tests may still be worth the price, for harsh debates may prove fruitful.⁶⁸

The knowledge argument, then, may lend some weight to the case for open testing, but it is a weight of uncertain heft.

3.4 The Learning Argument

According to this argument, open testing could increase comprehension of the subject/matter on the part of those tested by providing for feedback from the test. However, test publishers reply that any presumed learning effects in this context are illusory. ACT states:

Norm-referenced tests, such as those used in the ACT battery, are developed to aid colleges and universities in making decisions related to educational plans, admissions, and course placement and sectioning. They necessarily differ in their basic design from tests that are developed to guide instructional decisions and facilitate student learning. This latter purpose is best served by criterion-referenced tests constructed to yield scores that are meaningful in a particular instructional context.

To use the ACT tests in an instructional context . . . is to invite if not prescribe a serious misuse of these tests.⁶⁹

Wellington Grimes of the National Association of Independent Schools suggests, moreover, that in general ". . . the value of using test results for teaching purposes is an argument observed more in theory than in practice."⁷⁰

ACT has offered an even more elaborate and convincing defense to the learning argument:

Tests of the type contained in the ACT Assessment Program are designed to sample from a broad domain of knowledge and intellectual skills acquired and developed throughout the students' educational experience. They are not designed to measure whether specific facts have been acquired or specific instructional objectives have been met. The emphasis is on whether students can generalize from such facts and apply the principles in the context of the college curriculum.

As a consequence, access to tests and test answers may result in students knowing the correct answer to a specific test item. It will not, however, lead to students' acquisition of the general skills that the individual item represents. There is simply no effective short-cut to learning these higher-order skills. They can best be acquired through sustained instruction and serious study throughout the secondary school years.

To imply . . . that making the tests and answers available to students will contribute to their educational growth is at best optimistic and at worst misleading and irresponsible. The problem is exacerbated if the interpretation of the test items and performances is left to individual students.⁷¹

Open testers rejoin that knowledge of the corrected results, coupled with repeated exposure to the tests, at least teaches students how to take standardized tests, which is itself a form of learning. This idea has an element of plausibility, but belongs in the realm of opinion. We will return to this question when considering the issue of teaching to the test; for now, suffice it to say that the argument on the possible educational benefits of open testing is strong on opinion and slim on fact -- at least with respect to standardized multiple-choice admissions-type tests. Such tests are not designed specifically for instructional purposes, nor are they geared to particular curricula.

Let us now turn to the opposing arguments, those for secure testing.

3.5 The Test Quality Argument

Secure testers have raised concerns about test quality. These concerns have dealt mainly with test validity and test equating.

Assertions that invalidity would result from question re-use in open testing deserve close attention. Secure testers seem convinced that this is so, and no analyses of experiments with question re-use seem to exist. Open testers think that if enough questions and answers are made public, standardized tests can still be valid even if questions are re-used. For example, they would argue, if every SAT ever given were published, you'd have over 50 years of questions -- literally tens of thousands -- out in the open. As a practical matter, a prospective test taker would be foolish to try memorizing thousands of questions and answers rather than mastering the substantive basis of the tests.

One way of analyzing this idea is to examine two ratios. The first ratio may be thought of as the total number of re-used questions on a particular test (NR), divided by the total number of published questions from all previous tests (PT), or NR/PT . The second ratio may be thought of as the total number of re-used questions on a particular test (NR), divided by the total number of questions on that test (TT). It can be postulated that as long as both ratios NR/PT and NR/TT are small, validity will not be adversely affected. While I know of no studies dealing with this postulate, it has a strong intuitive appeal. It stems from the common-sense reasoning that a small NR/PT ratio suggests a small benefit from memorizing all the previously asked questions to begin with, and that a small NR/TT ratio means that whatever benefit exists will have a minimal effect

on a person's test score. This "ratio approach" deserves further investigation by researchers in the testing field.

What is one to make of this? We have two appeals to common sense, without any hard data to support one version on the consequences for test validity over the other. Test re-use thus seems to be a useful subject for research by open and secure testers alike.

Some current instances of re-use shed a very hazy light upon the area. State bar exams are commonly made public after administration, and questions are thought to have been re-used without dire results. The National Assessment of Educational Progress (NAEP) makes its questions public, and while it does not re-use them on its own tests, some states pick up the "discarded" questions and use them on state assessment tests.⁷² The state of Michigan has gone in for a different sort of re-use: it has given the same statewide assessment test for two consecutive years. Michigan seems satisfied with the results; in the words of one official, if publication or question re-use per se lead to student knowledge, you couldn't prove it by Michigan, since scores have been drifting downward.⁷³

These examples of question re-use are far from determinative. The portions of state bar examinations that are released to the public, for instance, differ from classic secure standardized tests in that they are largely essays, while the latter are multiple-choice in format. Since essays are free-form, revealing essay questions may pose a smaller threat to question validity than revealing multiple choice questions. And since Michigan's state tests are not used in any manner that has direct consequences for individual students, they have little incentive to read the questions.

Open testers might ask whether test publishers who re-use their own secure questions over time do not already run the risk that some test takers have been exposed to these questions before. This could occur if a person took a secure test such as the SAT several times, not during back-to-back administrations, but spread over a longer period. How frequently this occurs, if at all, is unknown. If this limited "prior exposure" were of major concern to publishers, they would not re-use even secure questions.⁷⁴

The assertion of secure testers that the number of good questions is finite deserves examination. In certain fields the array of questions, while not infinite, may be quite vast. A math test, to give just a simple example, might measure skill in multiplying two digit whole numbers. In our number system there are 90 of these (10-99); thus, there are 90×90 , or 8100, questions available for measuring this skill. This number would increase many times with different formats in which the questions might be posed (e.g., written horizontally or vertically). In any case, while the number of possible questions is indeed finite, it is still very large. Hence it seems implausible that a student faced with a test of this skill would try to memorize 8100 answers; it is far more likely that he or she would practice the skill itself.⁷⁵

Similarly, tests of reading comprehension could be almost infinitely varied by changing the reading selections (even while holding the level of difficulty constant) and the particular questions, not to mention the shadings of the correct and alternative answers provided.⁷⁶ The point open testers might make here is that students faced with a large, though finite, number of questions will find it more efficient to study the skill

those questions measure than to memorize, or cram, answers to questions that might be on a test. When students know the basis on which they are to be judged, and go about improving themselves on that basis, that state of affairs is not one to which most people would object.⁷⁷

To the argument that open testing must inevitably lead to test invalidity, Walt Haney has the following reply:

Such flat assertions of item invalidity as necessarily resulting from open testing are simply erroneous. This is clear from the fact that several sorts of test questions are currently publicly available and still can be validly used - essay test questions, and publicly released items from National Assessment now being used in several state assessment programs, are just two examples. As a result, it is not appropriate to view the issue of test validation as an all or nothing issue, but rather to think in terms of the conditions under which, and the degree to which, public release would result in diminished validity.⁷⁸

Haney goes on to mention

. . . three possible mechanisms for developing tests which would allow for release of corrected tests after administration, and would at the same time minimize threats to test validity.

1. Reliance on test items which directly measure relevant skills, essay questions for example, rather than indirect indicators of real skills as in the case with multiple choice items;

2. Development of item generating algorithms, which could produce large numbers of items equivalent in terms of content;

3. Development of large test item banks from which relatively small samples of items could systematically be drawn so as to yield alternative test forms.⁷⁹

Implicit in such suggestions, however, are some very different notions of test validity and equating than presently relied upon by test publishers. The form of validity evidence currently relied upon by developers of secure selection tests is predictive validity. The tests

are validated by the subsequent performance of applicants admitted to relevant institutions. The SAT, for example, is most frequently validated in terms of its power to help predict students' college grades. Suggestions such as (2) above for relying more heavily on content validation of tests may indeed be feasible, but may carry a cost in diminished predictive validity. To make this point more clear, let us go back to the two-digit multiplication example. While both of the problems 10×10 and 27×89 are examples of two-digit multiplication, the former is far simpler than the latter. Hence the latter likely would have more power to discriminate among skill levels. Thus, heavier reliance on content validation and equating may indeed diminish predictive validity or at least complicate efforts to maintain such validity in admissions tests. The point which should be made clear, however, is that validity resides not in a test per se, but rather in the way it is used and in the inferences drawn from test results.⁸⁰ Since we simply do not have much good information on exactly how currently secure admission tests actually are used, and no evidence at all on how open admission tests would be used, the debate regarding validity consequences must remain largely speculative.

Yet it is not at all clear that open testing necessarily would require new methods of validation and equating. For example, items used strictly for pretesting and equating could be kept secure and not be counted toward an individual's score vis-a-vis contemporary test takers (as is sometimes done at present). This approach has

been taken in the New York law. Secure pretesting and equating questions could be maintained, while test items that directly affect individuals' scores are to be released.⁸¹

The question nevertheless remains whether a system requiring strict segregation of secure items for pretesting and equating from open items contributing to test takers' raw scores would lead to more complex and costly testing programs.⁸² This brings us to the second of the arguments in favor of maintaining secure testing.

3.6 The Cost Argument

Secure testers have made the point that if open testing is required publishers will be forced to use expensive alternative procedures for validating and equating tests. The dispute over exactly what new techniques would be required -- e.g. would all-new test forms be needed or could some questions be re-used? -- gives this question a certain elusiveness. Moreover, as secure testers have argued, other real costs may be associated with open testing (fewer administrations, possibly lowered quality) that are not readily converted into dollar figures. Granted this fundamental ambiguity, the estimates advanced so far give at least a rough idea of the cost implications of open testing.

The incremental cost estimates of open testing cited by test publishers -- \$4 or higher per test taken for the SAT and \$25 or higher for the MCAT -- suggest that costs of open testing would vary with the number of test takers over whom development costs could be spread. The proposed "Testing Act of 1979" (H.R. 4949) implicitly recognizes this in requiring that only tests administered to more than 5000 people

annually be open.

Open testers suggest that publishers' estimates are too high. NYPIRG, for example, argued in a memorandum supporting the New York bill that "The cost of developing new test questions is a minimal percentage of the total cost to the test-taker."⁸³ ETS critic Allan Nairn has asserted that "By ETS's own figures, they spent 5 percent of test revenues on question development, but make 22 to 27 percent profit."⁸⁴ In support of that statement, Nairn has cited an internal ETS document, "ETS Activity Analysis for 1967-1971," comparing test development costs with total test costs. If open testing affects only development cost -- since overhead and administrative expenses need not vary between open and secure tests -- that ETS document indeed suggests that development cost amounts to roughly 5 to 10 percent of total costs.⁸⁵ ETS figures for 1970-71 for some selected tests, based on dollars per session per candidate, are as follows:

	<u>Development Cost (DE)</u>	<u>Total Cost (TC)</u>	<u>%DC/TC</u>
SAT	.32	4.80	7%
LSAT	.49	10.83	5%
NTE	1.01	6.34	16%
GRE	.62	7.17	9%
PSAT	.01	.54	2%

These figures prove nothing for certain. It is possible, for example, that the costs cited for 1970-71 are rock-bottom because of question re-use. Nonetheless these figures may be useful in estimating the increases in test cost likely to result from open testing.⁸⁶ For instance, the \$.32 development cost for the 1970-71 SAT would have to be multiplied over 12 times to approximate CEEB's threatened \$4 increase.

Such costs, of whatever magnitude, must be compared with the supposed advantages of open testing. People who view secure testing as absolutely preferable to open testing would regard even \$.32 as too high a price. Those who view open testing as socially beneficial might be willing to absorb a \$4 increase or more in the cost of SATs. The exact balance of cost and benefit remains to be determined, but the cost estimates of some secure testers seem greatly overstated in light of available facts. Should publishers have more concrete data to refute this analysis, that information would be most welcome, but in its absence the benefit of the doubt seems to swing to the open testers.⁸⁷ As one observer has noted, " . . . it is impossible to judge this issue until test publishers provide us with a detailed accounting of what current test development, administration and scoring costs are."⁸⁸ In sum, then, some tangible rise in cost is likely to be associated with open testing, but available evidence suggests it need not be great.

3.7 The Test Preparation Argument

The essence of this argument is that open testing would lead to an overemphasis on testing as students and teachers scramble to study disclosed test items rather than concentrating on substantive learning. It is far from clear, however, that open testing would leave us vulnerable to such

concerns, or that it would make things worse than they are.

To begin with, the supposed bonanza for commercial review courses must be compared to the already formidable size of this industry. According to The New York Times, the coaching industry has been estimated to be a \$60-million-a-year business. Stanley Kaplan, owner of the largest chain of test review centers, operates 80 centers and says this number has quadrupled since 1970.⁸⁹ With millions of standardized tests given yearly in this country, preparation is already big business. Bar review courses, which are taken by virtually all law school graduates, cost from \$225 to \$600 per student, while SAT review courses usually go for \$25 to \$300.⁹⁰ Advocates of open testing argue that this state of affairs already has created the bonanza warned against, and that open testing will reduce the need for coaching schools since students could prepare for tests on their own once test content is disclosed. Marjorie Dunbar, education chairman for the New York chapter of the National Association for the Advancement of Colored People, suggests:

Many students cannot afford these expensive coaching courses and are therefore denied equal access and opportunity. This bill [the NY law] would enable all students to get access to post-test questions, not just those who can afford the courses offered by 'cram' schools.⁹¹

Other observers have advanced similar arguments, namely that the present commercial test-preparation services embody real inequities. Former ETS researcher Lewis Pike has been quoted as saying, in reference to test-coaching schools, "The rich, in terms of test scores, were getting richer, at the poor's expense."⁹² Such concerns have been heightened in recent months by the findings of a staff report released by the Federal Trade Commission showing that some students, particularly underachievers, may

be helped by coaching.⁹³ If this is so, then open testing might beneficially increase access to test preparation. Rather than paying for the expertise of coaching schools, the poor and the motivated could help themselves or get less expensive tutoring based upon released test items. Others have suggested that currently advantaged students would instead reap the bulk of any gain from released test items and thus further widen the gap between the affluent and the poor. David McClelland, Professor in Harvard's Department of Psychology and Social Relations, for example, has suggested that it is difficult in any system of assessment to prevent people who are strongly self-interested from gaining unfair advantage.⁹⁴

To sum up, two scenarios have been suggested regarding the impact of open testing on inequities in test preparation and coaching. The NAACP support for open testing and the FTC report suggest that underachievers have the most to gain from coaching. If they were to gain similarly from open testing, inequities in test preparation might be reduced. On the other hand, others have suggested that open testing might do little to reduce inequities and could increase them. The method of implementing open testing would likely condition which scenario proved out. If past test questions were widely disseminated, for example through routine return of corrected test results to all test takers, reduction of inequities in test preparation would seem more likely. If, on the other hand, past questions received only limited distribution, for example if individuals had to pay substantially more to acquire corrected results, then inequities might continue or even increase. Whatever the case, if we are serious about reducing inequalities among students in preparation for selection tests, avenues of attack more direct than open testing alone would seem more promising.

In short, arguments concerning implications of open testing for test preparation are far from conclusive. Similarly, most of the arguments previously reviewed cannot be proved definitively. Before trying to pull these arguments together in any kind of summary or conclusion, let us step back in Section IV to review some of the broader issues which are likely to influence the outcome of public debate on open testing.

PART IV: BROADER PERSPECTIVES

The resolution of the dispute between open testing advocates and secure testing defenders won't hinge solely upon the arguments of adherents to either position. Political, judicial, and situational factors also play a part. While it is impossible in this brief review to explore all these perspectives fully, some mention of them will help clarify the broader context of the debate.

Political Factors. Since much of the impetus for the current debate over secure versus open testing derives from legislative proposals, it is not surprising that several different political themes are reflected in the debate. Prominent among them are concerns for interest group politics and government intervention.

Concerning interest group politics, it has been suggested that the debate over open testing is but one skirmish in a larger battle between test critics and enthusiasts. On one hand, open testing may be viewed as powered largely by the liberal education establishment's opposition to testing generally. Objective tests appear to show that the quality of education and learning in the U.S. is decreasing. CEEB and ETS, for example, reported that SAT scores continued to decline in 1979.⁹⁵ Instead of accepting this sorry message, some members of the education establishment may have set out to attack its bearers. One essayist -- interestingly a public school teacher -- has written:

The explosive growth in the education Establishment has been, and is being, drawn from among the weakest of our college graduates. It is therefore entirely consistent that they attack or drop IQ testing, ability grouping and objective tests for teachers and administrators. . . .⁹⁶

On the other hand, some opposition to open testing may stem not simply from public policy considerations, but rather from a desire to defend the testing establishment generally. Test publishers and sponsors, for example, may simply be trying to defend their profession and reputations from what they see as a wider onslaught by ill-informed test critics.

A second political theme reflected in the debate is increasing government interference. Why should government tell private testing agencies how to run their affairs? As one observer suggested in legislative hearings in Congress on September 10, 1979, why should the U.S. Congress require open testing by private testing agencies, when the federal government itself, as the sponsor of civil service exams, is one of the largest secure testing agencies in the country?⁹⁷

While these themes are important, they are only part of the political story behind the open testing debate. Legislative proposals for open testing have been promoted not just by elements of the education establishment, but also by student and public interest groups. Also, the Commissioner of Education in New York, Gordon Ambach, surely a member of any education establishment elite, actually testified against the New York open testing legislation. Moreover, some testing professionals, such as Thomas White, do seem at least somewhat sympathetic to the idea of open testing. And though the current legislative proposals in Congress seem to be an instance

of the federal do-as-I-say-not-as-I-do proclivity, surely it is not an unusual one. The Congress, for example, continues to exempt itself from the affirmative action guidelines required of private industry. Perhaps the ultimate evidence that current concern over standardized testing cannot be explained by simple politics is that one of the testing reform proposals now before the House of Representatives was introduced by a conservative Southerner (Sam Gibbons of Florida), and the other by a liberal Northerner (Theodore Weiss of New York).

Judicial Factors. The contrast between the private and public roles of secure standardized tests suggests that the courts may well influence the outcome of the debate. To what extent, if at all, do the proprietary rights of test publishers conflict with the disclosure requirements of open testing legislation? Under what conditions might freedom-of-information laws extend to contents of tests serving as instruments of public policy? It is entirely possible that the courts may play a major role in the resolution of such questions.

The recent surge in litigation has not left the educational sphere untouched, and testing has attracted its share of lawsuits. In the words of Melvin Novick, ". . . it is clear that the content of and manner in which information is transmitted from a testing organization to an educational institution will be the subject of close legal scrutiny."⁹⁸ Indeed there is ample precedent to show that the courts will entertain modification of current testing practices. As Novick notes, "In the areas of due process and consumer rights, the public is restless and much less willing to accept even the appearance of arbitrary administrative treatment.

DeFunis, Bakke, and K.D. challenged the system on important issues and there will surely be others."⁹⁹ Recent challenges to secure standardized tests have raised the possibility of a test case where a student or parent demands the release of test questions and answers after test administration.¹⁰⁰ Corbett writes: "Perhaps the issue of openness in testing could be resolved by a parent bringing suit to see any test given his/her child and the responses to it. To the best of my knowledge this has never been done and yet it seems so simple."¹⁰¹

Procedural remedies are especially dear to the hearts of most lawyers, for they have one overwhelmingly appealing virtue: the substantive complexity at the heart of a given controversy need not be dealt with. If open testing were to be widely perceived as a procedural safeguard against abuses in testing, then its chances would rise spectacularly. Under open testing, such a scenario would suggest that interested parties would examine tests with a fine-tooth comb, and misclassification and the like would end. As with Adam Smith's "invisible hand," the self-interests of the tested (and their parents) in more accurate tests would redound to society's benefit. Secure testers could reply with some justification that such an ideal outcome is illusory -- that the courts would become unbearably crowded with the just and unjust claims of a horde of test takers, all intent upon having scores adjusted in their favor.¹⁰²

How much the courts might be drawn into such disputes is unclear. There is a fairly long history of court cases dealing with employment testing, but not with educational testing generally or testing for admission to educational institutions.¹⁰³ It may well be that without clear precedent or specific legal grounds, the courts will tend to steer clear of disputes over testing.

Situational Factors. The desirability of open versus secure tests is not an all-or-nothing proposition, but rather is situation-specific. A number of considerations might tend to strengthen or weaken the case for open testing. Some of these are discussed below.

1) Voluntary tests. If the examinee is taking a test of his or her own volition, then the security of the test should be of minimal concern. However, to the extent that there is direct coercion (an institution that administers secure tests to all pupils as a matter of policy) or indirect coercion (a student is free to refrain from taking the LSAT, but only at the cost of not being considered for admission by most American law schools), the fairness claim in support of open testing would appear stronger. In this respect most of the secure tests so far mentioned in the debate do contain an element of coercion.

2) Public or private sponsorship. The norms of our society are that public institutions are held to stricter standards regarding public disclosure of information than are private institutions. Thus one could argue that a public institution (such as a state university) should be most restrained in the use of secure tests while a private institution (a private college or research institution) should be allowed more latitude.

3) Consequences for the tested. If a test is used for group or program evaluation, with no clear consequences for individual test takers, then some of the arguments for open testing would appear to be less relevant. While such tests may have important effects (such as the NAEP program, or a school district testing itself for overall performance), they would have

only minimal impact on any given students whose welfare would not be directly at stake if test questions were mis-scored or ambiguous.¹⁰⁴

However, when tests are used for selection or admissions, the consequences for individuals could be direct, and open testing would seem more appropriate. The tests at issue in the debate over testing legislation are those which apparently have just such direct consequences.

4) Timing of test disclosure. Few advocates of open testing argue that the questions and answers of a test should be disclosed before administration. However, once the test has been given the interest of the public (and the examinee) in knowing the test contents would seem to become stronger vis-a-vis the interest of the test sponsor (and the public) in preventing cheating on the test. As time after administration lengthens, so the arguments in favor of security would seem to diminish. Indeed, at some point (say 100 years later), any security interest of the test publisher in re-using questions would become totally attenuated; but unfortunately, so does any benefit that might accrue to the examinee from making the test items known.

5) The size and impact of any added cost. If secure testers are correct in saying that open testing will boost test charges, it is highly relevant to know the size of those increases and upon whose shoulders they will fall. Some argue that students will end up footing the bill, and that impoverished students will be hit hardest unless we set up mechanisms to subsidize them. Winsor Lott, chief of the Bureau of Elementary and Secondary Testing for the New York State Department of Education, argues:

Our chief argument is that the inevitable increase in cost that would result from making these tests 'open' must eventually be borne by the candidates themselves. The population served by the programs that use 'secure' tests are those that can least afford the added cost.¹⁰⁵

In a similar vein, ETS states:

Lower income families would be hardest hit by these increases, and could find it very difficult to take advantage of special educational opportunities offered through testing--such as placement or credit-by-examination. Although "fee waivers" are now (and have been) available for students with limited financial means, if the (California) bill passes some very costly procedural and administrative requirements could reduce or eliminate these "fee waiver" funds.¹⁰⁶

If significant cost increases must result from open testing, then this deserves attention. Perhaps secure testing should be maintained for tests given to small numbers of people, where disclosure would sharply increase costs. Or perhaps means could be devised to subsidize costs for particular types of tests or test takers.

As this brief situational analysis suggests, alternative remedies should be fashioned to deal with various testing circumstances. Although a scheme of alternative approaches would probably be more administratively complex and difficult to interpret than either maintaining the status quo or mandating "pure" open testing, it might also be more in accord with the concerns underlying the debate over test security.

* * * * *

Before leaving this review of broader perspectives on the open versus secure testing debate, note that wide though the debate -- and the considerations bearing on it--may be, some observers have suggested that it is still too narrow--or at least focused on the wrong issue.

David McClelland, for example, points out that the particular nature of the secure selection tests is the real cause of much of the controversy over

these tests. In particular, he notes that security problems are connected with these tests partly because they are only indirect and respondent indicators of the real skills or attributes of interest. McClelland suggests that fundamentally different forms of assessment--direct as opposed to indirect, and operant rather than respondent--would do away with many of the problems of test security. Direct competency tests, for example, would directly sample the real skills of interest and hence would help remove the security problem. Indicator tests (such as the SAT) measure only indirectly, and hence one may challenge the worth of indicators (for example, that they are culturally biased, or invalid). But with direct competence measures--typing tests, or shooting baskets from the foul line, or the performance parts of driving tests--issues of bias and security rarely arise, even when some test takers pay for the advantage of drivers' test coaching, because the skills of interest are assessed directly.

Similarly, operant as opposed to respondent assessment would help do away with problems of test security and bias. In operant assessment, test takers actually have to construct an answer or perform an action rather than merely choosing among preselected alternative answers. With operant assessment, security is less a problem because it is impossible for a test-taker simply to memorize a sequence of answers--or a test-taking technique--without knowing or practicing the underlying competency that the test is designed to measure.¹⁰⁷

Observers of the broader issues in testing and assessment readily acknowledge that their proposals by no means address all of the issues raised in the debate. Such basic issues concerning the nature of

assessment clearly predate the open testing controversy. McClelland, for example, has advocated direct competence assessment for more than two decades. Yet if this controversy has done nothing else, it seems to have stimulated greater attention to such fundamental issues.

PART V: CONCLUSION

Drawing conclusions on the merits of open versus secure testing is by no means easy. Opposite sides in the debate base their arguments largely on different premises, and on many key issues there is little evidence available on which to base a judgment. Moreover, a welter of wider considerations, only some of which have been touched upon here, clearly color perspectives on the issues under debate. One such consideration is historical.

Admission testing practices have not been static historically. One of the major changes relevant to our discussion occurred in 1958: that was when students were first allowed access to their own SAT scores. Looking back from 1979, when the debate is focused on release of test questions, it is hard to imagine a time when test takers were denied not only the basis for their SAT scores, but the very scores themselves.

Interpretations on the 1958 change vary. The College Board writes: "To serve an ever-widening student clientele and a rapidly diversifying membership, the Board's information services to students included (for the first time in 1958) the release of scores and interpretive information to students" ¹⁰⁸ Previously, only the schools designated by the candidate had received the SAT score. Daniel Boorstin writes of the new policy: "This was, of course, a significant new step in reinforcing public awareness of statistical communities of the mind. The president of the College Entrance Examination Board, attacking the taboo of silence, boasted now that the publicity for test scores was 'a triumph of morality.'" ¹⁰⁹

A more detailed account of the events of 1958 has been provided by Michael Schudson:

In 1958 the Board authorized high schools -- for the first time -- to tell students their scores on Board exams and to interpret the scores for the students. Nothing could be more indicative of the extent to which the Board had been the servant of the colleges -- the educational system, if you will. Board executives sought to shift this long-standing preoccupation with the colleges' needs to give some attention to the applicants' interests. The decision to release the scores was regarded by Frank Bowles, President of the Board from 1948 to 1973, as the most important act of the Board in those years. The act was a symbol of the increasing importance of choosing a college and being chosen by one and was in part an unintended cause of the growing anxiety over test scores among the college-bound and their families.¹¹⁰

"Unintended consequences" often accompany reforms. This does not mean reforms are not worth while, but only that reformers must be aware of possible side effects of their proposals. Open testers should seriously consider such possibilities -- for example, that increased anxiety over testing or "testolatry" might be a spin-off of open testing.

A second general comment before making any evaluation of security in testing is that both sides of the cost-benefit equation must be considered. Advocates tend to point to the benefits of their system and the costs of their opponents' approach. While that might be expected from opposite sides in a debate, it must be recognized that both open and secure testing schemes appear to offer benefits at certain costs. It is by the overall balance of benefits versus costs that a testing system should be weighed. Bearing this in mind, the argument of CEEB might be considered:

College Board programs have pioneered in efforts to eliminate quotas, favoritism and personal bias as barriers to educational and economic opportunity. National standardized tests sponsored by the College

Board open doors and expand opportunities and reduce the importance of more subjective, social, personal and economic factors in determining access to post-secondary education. Impairment of the standardization, integrity, and availability of such tests, would actually decrease educational opportunity.¹¹¹

If CEEB's assessment is correct, and if "impairment" were an inevitable consequence of open testing, then secure testers would be in a formidable position. The point is that the overall net gain or loss from open testing should be the bottom line for decision, rather than any individual line item.

Third, one's attitude toward open or secure testing may be largely shaped by individual definition of the goals of measurement and the uses to which test results are put. Shlien, for example, writes:

Is the elimination of academic failures worth the price of excluding a good many who could well succeed, given the chance? . . . Is exclusion by test any more unfair or damaging than exclusion by other misfortunes, such as lack of money? Hard to say, but it seems that the test speaks with more finality. One can earn money, but the test tends to assign a fixed status, hard to change, seemingly beyond redress for most.¹¹²

If open testing uncovers a bias that penalized 1 in 1000 or 100 in that 1000, we still need to determine how to rectify that bias and what costs we are willing to absorb to do so. If secure testing allows us to maintain our present selection system, which works relatively efficiently on the surface whatever its underlying flaws, are we willing to maintain that system although individual test takers are denied access to corrected test results? If we could quantify the relative costs and benefits of open versus secure tests, the question would still remain of how we choose to weigh factors and what we as a society value enough to pay for. For

example, if the trade-off boiled down to efficiency versus openness, which would we choose? Would we accept efficiency as long as only 2% of students were treated unfairly? Or 20%? Or .2%?

Fourth, open testers and secure testers, by and large, have not been arguing about the same things. Open testers concentrate on public policy considerations while secure testers concentrate on the mechanics and techniques of testing. This is suggested indirectly in the memorandum from Thomas White mentioned earlier:

. . . I have urged the ETS officers to separate the arguments of convenience, efficiency and cost benefit from those of technical impossibility. I know that given a reasonable lead time, a totally "open" testing and admission processing system for legal education is technically and economically feasible. I expect that some reduction in convenience (service) and some increase in cost would result from required changes, but there is no reason why it cannot be done.¹¹³

This double focus to the debate is another factor that makes it difficult to reach conclusions. On many basic issues, arguments by opposing parties rarely meet head on. Yet the two sets of issues -- the broader public policy concerns raised mainly by open testers and the narrower test quality and cost considerations generally focused upon by secure testers -- ought to be addressed on their own merits.

Finally, secure testers are not on firm ground when they argue that disclosure is already sufficient to meet all the reasonable objectives of open testers. ETS, for example, says:

But for all that, the tests are not secret . . . All programs with which we are familiar will make sample questions and directions available. In addition, actual tests for virtually all national testing programs are available for professional scrutiny and have been for many years. The Mental Measurement Yearbook is probably the most

famous example of such external professional review, which typically includes scrutiny of test questions against specifications and statistical reports.¹¹⁴

Yet, in a speech delivered at the University of Iowa in March 1977, Oscar Buros, editor of the Mental Measurements Yearbook, put forth a quite contrary view. He declared: " . . . the information available to permit an adequate assessment to be made of these secure tests is quite unsatisfactory. Although our reviewers generally receive some in-house material, which is not available to other educators and psychologists, even this material is inadequate."¹¹⁵ Buros went on to say:

I would like to repeat a statement which I made forty-two years ago:

Today it is practically impossible for a competent test technician or test consumer to make a thorough appraisal of the construction, validation, and use of most standardized tests being published because of the limited amount of trustworthy information supplied by test publishers and authors . . . If testing is to be of maximum value to schools, test authors and publishers must give more adequate information . . . It would be advantageous . . . if test publishers would construct only one-fourth to one-half as many tests . . . and use the time saved for presenting the detailed information needed by test consumers.

Unfortunately, although some progress has been made, my 1935 complaint is equally applicable today to the majority of existing tests -- and especially so for secure tests.¹¹⁶

In short, the assertion that current review and disclosure procedures are sufficient to guarantee high quality tests is challenged not only by open testing advocates, but also by the very author whose work has been cited to bolster the assertion. At the very least, this indicates that the adequacy of current public information on secure tests is a legitimate point of discussion.

I have wrestled with these questions in reaching my own conclusions. I should reiterate here that these are my own tentative conclusions and not necessarily those of others affiliated with the National Consortium on Testing or the Huron Institute.

First, I must acknowledge that open testing may have consequences that are unanticipated and undesirable -- for instance, a sharp increase in the complexity and cost of selection procedures, or large cutbacks in standardized testing for these purposes. This could happen if open testing led to unwarranted attention to, and legal challenge of, specific test questions regardless of overall test quality. Maybe Stanley Kaplan is right in saying that testing is a Pandora's box, and that opening-up secure tests would unleash forces that would tear all tests apart and push us back to the evils of favoritism in selection procedures.¹¹⁷ Maybe open testing would exacerbate the tendency to focus measurement on that which is most easily measurable (arithmetic, vocabulary) and neglect skills that are less amenable to measurement (creativity, problem-solving). Maybe it all comes down to whether we take confessedly unfair actions from time to time in using imperfect secure tests in selection to avoid the paralysis of insistence upon unattainable levels of fairness in life. Maybo.

For all that, to me the balance of the arguments reviewed points to the conclusion that, in the public interest, the current status of secure testing needs to be changed. First, I subscribe to the view that in our society the burden of proof for security -- or secrecy, as it might be put -- in public practices must rest with those who advocate security or secrecy.¹¹⁸ Second, I find the arguments for open testing marginally more persuasive than the rebuttals. Open testing would enhance the goals of fairness and accountability (and, to some extent, knowledge about the tests), though open testing per se seems unlikely to have much educational

value. In short, open testers seem to have made a sufficiently convincing plaintiff's case that the burden of proof initially placed on them moves to the defenders of secure tests.

This burden of proof has not, in my opinion, been met by the generally weaker arguments advanced in support of secure testing. Statements that invalidity must result from disclosure, and that present disclosure is adequate, do not hold water. Costs may increase somewhat under open testing, though CEEB estimates seem greatly inflated in light of the skimpy supporting data available. In disputes of this sort my inclination is to say that if party A (in this instance the test sponsors and publishers) has facts and figures that would prove A's contention, but does not release them, judgment must be for the opposing party. Finally, although I don't believe open testing will eliminate all the problems associated with test coaching, neither do I believe that open testing is likely to make the existing state of affairs significantly worse.

Others have reached their conclusions about open testing on somewhat different grounds. Haney has stated:

My own summary view is that if one accepts the validity of the proposition that individual test-takers have a right to know the basis on which they are judged -- and I would like to make clear that personally I do -- then issues of test development and technical quality need not necessarily stand in the way of open testing policies designed to implement that right.¹¹⁹

If open testing is a worthwhile goal, one still can question whether legislation is a desirable means of pursuing it. After all, many advocates of open testing grant that legislation is not their preferred method of effecting change. It is worth noting that the 1958 CEEB decision to return SAT scores to test takers was due not to any legislated reform, but to internal Board policy. Establishing open

testing policy through public law may result in diminished flexibility to meet other goals of assessment. Moreover, it might not even lead to complete elimination of the "secrecy" and nonaccountability in testing that open testers oppose. It is clear, for instance, that some failures to return corrected results to test takers do not stem from explicit policies of test publishers or sponsors. Beck and Stetz, for example, found in their survey of over 3000 teachers of grades K-12 that more than three quarters indicated that they did not report standardized test results to students.¹²⁰ This may have been for any number of reasons. Teachers may not have regarded the return of results as useful, or may never have been aware of the possibility of returning corrected results. Perhaps corrected results were not available in a reasonable period of time. This example is drawn from the realm of elementary and secondary school achievement testing rather than post-secondary admissions testing, but it does suggest some caution regarding the likely impact of legislation on testing practices. Specifically, some security or lack of openness in testing may stem not from explicit policies of testing agencies but from practical constraints or policies of educational institutions. Whether such constraints would be altered greatly by open testing legislation is clearly open to question. Many open testers readily acknowledge such uncertainty. They have turned to legislation as a means of testing reform not necessarily because they saw it as ideal, but because they felt unable to gain a hearing, much less significant action, on their concerns through other avenues.

However these concerns are met, resolution of the open testing debate

will depend largely on answers to a question which can be simply stated. What purposes and whose interests are to be served by standardized testing either in the schools or for postsecondary admissions and selection? If publicly mandated standardized testing is to directly serve the interests of those tested, on what grounds are we to argue that test takers should be denied access to the corrected results of the tests they are required to take? If the testing debate bears no other fruit, perhaps it will at least prompt closer consideration of the exact functions of these standardized testing programs and the rights and interests of those tested under them.

FOOTNOTES

1. Jon Haber. Testimony before the California Senate Education Committee, May 3, 1978. Reprinted by the Associated Students of the University of California, p. 1.
2. In addition to California and New York, "truth in testing" legislation has been introduced in the Texas and Maryland state legislatures, and similar proposals have been discussed in Colorado and Massachusetts.
3. Stephen Solomon (Director, NYPIRG Statewide Testing Project). Telephone interview, February 7, 1979.
4. Lawrence K. Altman, "Citing New Law, Medical Schools to Bar Entry Tests in New York," The New York Times, July 18, 1979, p. 1.
5. Transcripts of these hearings are not yet available. For an account of a less formal debate on open versus secure testing see W. Haney and A. Strenio, "Summary of the Spring 1979 Conference of the National Consortium on Testing, April 30 - May 1, 1979," Cambridge, MA: Huron Institute, 1979; and for an extended argument against open testing the paper entitled "Why Should Some Tests Be Secure," by John Fremer and Alice Irby, which is appended to this summary.
6. G. McDaniels, "Assessing Handicapped Students." In W. Schrader (ed.), Measurement and Educational Policy (San Francisco: Jossey Bass, 1979), pp. 3-7.
7. Ibid.
8. In a related area, on January 23, 1979, the Massachusetts State Board of Education adopted new regulations governing basic skills assessment (the Massachusetts version of competency testing) to go into effect in 1980. One provision of the regulations was largely the result of the initiative of William Corbett, an elementary school principal in the state. Under the terms of this provision, students and their parents will be guaranteed access to tests used in statewide basic skills assessments. While no copies of the test booklets will be allowed out of the schools, students and parents will have the right to review the questions and corrected answers in the company of school guidance personnel. This might be called a semi-open testing approach.
9. D. B. Guralnik (ed.), Webster's New World Dictionary, 2nd College edition, New York: World Publishing, 1970, p. 8.
10. The type of test involved, and the use for which it is designed, may affect one's conclusions regarding open testing. That is to say, domain-referenced testing, or diagnostic testing, may be different enough from norm-referenced multiple-choice selection tests to require independent analysis of the merits of open versus secure methodology.
11. The State of New York, Chapter 672 of the Laws of 1979, section 340 (1).

12. Edwin F. Taylor, "The Snares of Secrecy." Education Development Center News, Winter 1978, p. 1.
13. John M. Shlien, "Mental Tests and Modern Ethics." The Humanist, No. 6, 1958, p. 357.
14. William Corbett, "What Principals Can Do about the Secrecy Issue," Principal, January 1979, p. 59.
15. American Psychological Association. Ethical Standards of Psychologists. Washington, D.C.: APA, 1977. For more on this point, see W. Haney, "Standards for Tests and Test Use," National Consortium on Testing Staff Circular No. 3, Cambridge, MA: Huron Institute, October 1978, pp. 33-35.
16. California Senator John F. Dunlap. Press release dated May 3, 1978, p. 1.
17. See accounts of this incident in the following publications: Carey Winfrey, "Revised Law School Test Leads to a Dispute," The New York Times, April 28, 1978, p. A23; Steven Brill, "Law Boards Scandal," Esquire, April 25, 1978, p. 22; and Peter R. Melnick, "Changes in Law School Admission Test Increase Number of High Test Scores," The Harvard Crimson, April 13, 1978, p. 1.
18. Maxwell Gould, "MCAT Scores Fall from Past Level, Prompt Regrading," The Harvard Crimson, May 19, 1978, p. 1.
19. J. Wyatt Emmerick, "Error in GMAT Scores Affects 200 Applications," The Harvard Crimson, January 18, 1978, p. 1.
20. Kim Masters, "ETS's Star Chamber," The New Republic, February 5, 1977, pp. 13-14.
21. Corbett, op. cit., p. 58.
22. John D. Williams. Testing and the Testing Industry: A Third View, Grand Forks, North Dakota: University of North Dakota, 1975, p. 18.
23. New York Public Interest Research Group, Legislative Memorandum No. 25, undated, p. 1.
24. Taylor, op. cit., p. 1.
25. Jerrold Zacharias elaborates this point in his article, "The Trouble with IQ Tests," in Paul Houts (ed.), The Myth of Measurability, New York: Hart Publishing Company, 1977, pp. 77-78:

If not impossible, it is at least extremely inconvenient and time-consuming for someone without proper credentials to obtain full sets of aptitude tests or achievement tests. And to appreciate the implications of a test fully, it is necessary to study the full test, not just an occasional test item here or there.

26. NYPIRG Memorandum No. 25, op. cit., p. 2.
27. Haber, op. cit., p. 2.
28. Roger Shuy. Personal communication, January 9, 1979, p. 1.
29. Taylor, op. cit., p. 1.
30. The adoption of the new Massachusetts basic skills assessment regulations, referred to in footnote 8 above, may provide a precedent for more access to test results of this sort by parents.
31. Robert L. Ebel, "The Social Consequences of Educational Testing," in Anne Anastasi (ed.), Testing Problems in Perspective, Washington: American Council on Education, 1966, pp. 25-26.
32. Oscar K. Buros, "Fifty Years in Testing: Some Reminiscences, Criticisms, and Suggestions," Educational Researcher, July-August 1977, p. 14.
33. Stephen Solomon stated in a telephone interview with the author on April 17, 1979 that NYPIRG does not consider this a major argument for open testing and has not rested its position upon this particular ground. Gene Russianof, staff attorney to NYPIRG, reiterated this point in a telephone interview on April 21, 1979.
34. Mitchell Lazarus and Edwin F. Taylor, "They Don't Measure Learning," The New York Times, May 1, 1977, p. 22.
35. CEEB. Statement of the College Entrance Examination Board in Opposition to California Senate Bill No. 2005, May 11, 1978, p. 1. Cited afterward as CEEB/California.
36. CEEB. Memorandum in Opposition to Senate 5845-B by Mr. Lavalley, undated (prepared for use in the debate over the New York law), p. 2. Cited afterward as CEEB/NY.
37. Educational Testing Service. Memorandum of Opposition (to the New York bill), undated, p. 2. Cited afterward as ETS/NY.
38. Associated Medical Schools of New York. Memorandum in Opposition (to the New York bill), April 25, 1978, p. 2. Cited afterward as Associated Medical Schools/NY.

39. CEEB/NY, op. cit., p. 3.
40. CEEB. Memorandum for California Schools, Colleges, and Scholarship Sponsors about California Senate Bill [SB] 2005, May 10, 1978, p. 1.
41. ETS/NY, op. cit., pp. 1-2.
42. Ibid., p. 2.
43. Ibid., p. 1.
44. CEEB/NY, op. cit., p. 4 and CEEB/California, op. cit., p. 2.
45. Associated Medical Schools/NY, op. cit., pp. 1-2.
46. Ibid., p. 2.
47. Oluf M. Davidsen, President of ACT. Personal communication to California State Senator John F. Dunlap, May 2, 1978, p. 2. Cited afterward as ACT/Communication.
48. ETS (Western Regional Offices). Memorandum for Those Concerned with Standardized Tests and Testing in the State of California. May 9, 1978, p. 1. Cited afterward as ETS/California.
49. Dr. Richard Ferguson, Vice President, Research and Development Division of the American College Testing Program, expressed this sentiment in a telephone interview on April 18, 1979. He argued that it is better to concentrate on correcting individual identified problems with testing on a case-by-case basis rather than seeking a blanket solution through open testing.
50. ETS/NY, op. cit., pp. 4-5.
51. Mark G. Yudof. Memorandum to Walt Haney, May 17, 1979, p. 4.
52. At the same time, a test taker or a defendant might reasonably opt for an open test or hearing on the theory that any perceptible unfairness in treatment would be exposed to the public and hence made less likely.
53. Dr. Walter M. Haney. Testimony before the Committee on Education and Labor of the United States Congress, July 31, 1979, p. 3. Cited afterward as Haney/Testimony.
54. One example of this problem is given by M. Beck and J. Sklar, "An Assessment of Item Bias in the 1978 Metropolitan Achievement Test." Paper presented at the meeting of the Eastern Educational Research Association, Williamsburg, Virginia, March 1978. They reported a study in which three minority educators were asked to identify potentially biased test items. Although altogether they identified approximately 15% of the items reviewed as potentially biased there was little overlap in their identifications. In general, there is considerable disagreement as to exactly what bias means, both technically and in common usage. See, for example, R. Flaugher, "The Many Definitions of Test Bias," American Psychologist, July 1978, pp. 671-79.

55. Yudof, op. cit., p. 4.
56. Thomas W. White. Memorandum on "Truth in Testing," July 24, 1979, p. 1. This version of the memorandum is found in The Testing Digest, Volume I/ Issue 3, Fall 1979, p. 4.
57. Yudof, op. cit., p. 4.
58. ACT/Communication, op. cit., p. 1.
59. CEEB/California, op. cit., p. 3. However, one open testing advocate has proposed the following solution: "Prospectively, schools would have the option of having ETS conduct studies which would be subject to disclosure or of paying ETS for the raw data itself and then doing their own study (not subject to disclosure)." See Gene Russianof. Personal communication to Mike Rigas, December 7, 1978, pp. 3-4.
60. According to Robert J. Solomon, Executive Vice President of ETS, any improvement in spotting scoring errors will be minimal: "In 1967-77 we had requests to hand-score 1081 of the 120,000 law boards and found only four errors, all of them because students put the answers in the wrong spaces." See Edward B. Fiske, "Issue and Debate: Scholastic Testing Industry Facing Regulation by State," The New York Times, July 5, 1979, p. D6. Cited afterward as Fiske/Regulation.
61. The argument, then, is that open testing would serve to enfranchise the direct test consumers, i.e., the tested. It may well be simpler for publishers to serve but one master at a time (namely the sponsors); it is not necessarily desirable.
62. On the other hand, it is hard to see that nondisclosure of teacher-made tests would necessarily lead to improvement of their quality. Nevertheless, it is worth noting that not too many years ago standardized testing was seen as one remedy for the shortcomings of teachers' grading practices. See H. Kirschenbaum, et al., Wad-ja-get? The Grading Game in American Education, New York: Hart, 1971.
63. Fiske/Regulation, op. cit., p. D6.
64. Haney/Testimony, op. cit., p. 2.
65. Fiske/Regulation, op. cit., p. D6.
66. See W. Haney, "A Review of Publications on Testing for Parents and the Public," National Consortium on Testing Staff Circular No. 5, Cambridge, MA: Huron Institute, April 1979, for elaboration on this point.
67. For an account of past efforts to judge test quality, see W. Haney, "Testing the Tests," National Consortium on Testing Staff Circular No. 1, Cambridge, MA: Huron Institute, February 1978.

68. Yudof, op. cit., p. 5.
69. ACT/Communication, op. cit., p. 2.
70. Wellington Grimes. Personal communication, January 11, 1979, p. 2.
71. ACT/Communication, op. cit., p. 2.
72. According to Dr. Elinor Woods of the Huron Institute, fifteen states have testing programs which include some intact items from NAEP released exercises in tests for at least one area assessed. These states include Illinois, Louisiana, Maine, Massachusetts, and Washington.
73. Jack Schmidt (NAEP). Telephone interviews on January 11, 1979 and April 25, 1979.
74. See CEEB/NY, op. cit., p. 3: "... individual items or questions are reused and become the basis of 'equating' the scores from form to form and over time."
75. Walt Haney. Personal communication to Stephen Solomon, January 18, 1979, p. 3. Cited afterward as Haney/Solomon.
76. Dr. Ferguson said in a telephone interview on April 18, 1979 that the pursuit of "equivalence from test to test in content, scope, and character being measured [is] not an impossible task, but not an easy one either." He also expressed a tentative personal opinion that the conclusion stated in an earlier draft of this paper to the effect that the marginal direct costs of open testing are relatively minor is accurate. "The additional costs for us, given a large volume of test-takers, is not a staggering sum . . . However, there are hidden costs. It's sometimes hard to separate out these costs and their consequences aren't always that clear."
77. Haney/Solomon, op. cit., p. 3.
78. Haney/Testimony, op. cit., p. 4.
79. Ibid., pp. 4-5.
80. See American Psychological Association, American Educational Research Association, and National Council on Measurement in Education, Standards for Educational and Psychological Tests, prepared by a joint committee of the APA, AERA, and NCME. Frederick B. Davis, Chair. Washington, D.C.: American Psychological Association, 1974.

81. See section 342(1)(a) of the New York law.
82. See ETS/NY, op. cit., p. 2.
83. NYPIRG Memorandum No. 25, op. cit., p. 2.
84. Fiske/Regulation, op. cit., p. D6.
85. Herman F. Smith, Educational Testing Service, "Activity Analysis for 1967-1971," memorandum dated January 31, 1972. This document lists a number of categories as contributing to total cost. This summary of the ETS document treats all costs as either developmental or non-developmental. Listed as developmental are: Development of Measuring Instrument (Question Writing and Selection), Pretesting, and Item and Test Analysis. Listed as non-developmental are: Planning, Tooling-Up, Registration, Administration, Scoring Final Forms, Reporting Test Results, Maintenance of Score Report Files, Billing and Refunding, Transcript Processing, Special Studies, Operational Control, Security, Promotion and Circularization, Aids for the User, Custom Services, and Other.
86. Of course, these figures could be adjusted. Mike Garet asserted in a telephone interview on April 18, 1979 that the DC/TC ratio should take into account the number of items being re-used currently. Then, if under open testing no questions could be re-used, the development cost would go up additionally to reflect the need for more new questions. Allan Nairn, on the other hand, has asserted that the DC/TC ratio is misleading because it doesn't take into account the test publisher's charge to the public. In other words, the test cost to the publisher may be substantially less than the cost to the public (CTTP) for that test. Thus if one examines the ratio DC/CTTP, one might find an even smaller percentage than shown by DC/TC.
87. Several advocates of open testing have suggested that costs need not necessarily be born by test-takers in a single state, such as New York. They suggest that costs could be substantially reduced by spreading test development costs across test-takers nationwide.
88. Haney/Solomon, op. cit., p. 2.
89. Fred M. Hechinger, "Coaching Courses Studied by F.T.C.," The New York Times, December 19, 1978, p. C5.
90. Ibid.
91. Fiske/Regulation, op. cit., p. D6.
92. Steven Levy, "ETS and the 'Coaching' Coverup." New Jersey Monthly, March 1979, p. 52.

93. Gene I. Maeroff, "'Coaching' Findings May Change S.A.T.," The New York Times, June 5, 1979, p. C5.
94. See the Haney and Strenio, "Summary of the Spring 1979 Conference of the National Consortium on Testing," op. cit., p. 54. The author's personal experience with review courses and conversations with persons familiar with coaching practices suggest that some of these courses debrief their students after tests are given in order to obtain information about the items used and the wording of questions. This information is then passed on to subsequent purchasers of the coaching service. Data is not available on how widespread this practice is, but its existence suggests that unfair disparities in preparation can exist even with ostensibly secure tests.
95. Muriel Cohen, "Bay State S.A.T. Scores Drop for 6th Straight Year," The Boston Globe, September 11, 1979, p. 19.
96. Charles Frye, "Who Runs the Schools?," Newsweek, September 3, 1979, p. 13.
97. Haney/Testimony, op. cit., p. 2.
98. Melvin R. Novick, "The Influence of the Law on Professional Measurement Standards," an address presented at the ETS Invitational Conference on Testing Problems, October 29, 1977, p. 16 (Reprinted in ETS, Proceedings of the 1977 ETS Invitational Conference, Princeton, NJ: ETS, 1977).
99. Ibid., p. 19.
100. A detailed description of legal challenges to alleged test bias, test invalidity, and improper tracking based on test results can be found in Paul Weckstein, "Legal Challenges to Educational Testing Practices," Classification Materials, Cambridge, MA: Harvard Center for Law and Education, 1973, pp. 186-98 (1976 supplement, pp. 37-38). The article does not mention challenges to test security as such. However, resolution of some of these cases has clearly necessitated disclosure of test contents at least for the purposes of adjudication.
101. William Corbett. Personal communication, January 5, 1979, p. 1.
102. See John Feinstein and Jackson Diehl, "Question on P.G. Police Test Proves Difficult to Answer," The Washington Post, December 29, 1978, p. B1 for a sense of how this might occur.
103. See Summaries of Court Decisions on Employment Testing 1968-1977, New York: The Psychological Corporation, 1978.
104. This analysis, of course, takes a very narrow view of the potential impact of tests on students or education generally. To the extent that tests have any important consequences, direct or indirect, we ought to be concerned with their quality, including whether questions are misscored or ambiguous.

105. Winsor A. Lott. Personal communication, January 25, 1979, p. 2.
106. ETS/California, op. cit., p. 1.
107. For a summary of McClelland's views, see "Summary of the Spring 1979 Conference," op. cit., at pp. 50-54.
108. The College Board, "The First 75 Years," (Pamphlet published by CEEB), New York: CEEB, 1975, pp. 5-15.
109. Daniel J. Boorstin. The Americans: The Democratic Experience. New York: Random House, 1973, p. 225.
110. Michael S. Schudson, "Organizing the 'Meritocracy': A History of the College Entrance Examination Board," Harvard Educational Review, Vol.42, No. 1, February 1972, p. 62.
111. CEEB, California, op. cit., p. 3.
112. Shlien, op. cit., pp. 360-61.
113. Thomas White, op. cit., p. 1.
114. ETS/NY, op. cit., p. 2.
115. Buros, op. cit., p. 14.
116. Ibid.
117. This point was suggested by Stanley H. Kaplan. Telephone interview, February 1, 1979.
118. Haney/Solomon, op. cit., p. 2.
119. Haney/Testimony, op. cit., p. 6.
120. M. Beck and F. Stetz, "Teachers' Opinions of Standardized Test Use and Usefulness." Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1978, p. 10.